If you have any questions, Corbeil can answer them. I'm going hunting in the morning.

Abstract

Dr Bruce Wallace has proposed a regulatory gene model in which each gene in a linear array is preceded by a linear array of sensors, say A, B, C, etc., where A is the sensor for regulatory system A, B is the sensor for regulatory system B, and so on. When system A calls for the product of this gene then sensor A, and all sensors between A and the gene, and the gene itself must be reproduced; thus if there are many other sensors between A and the gene then reaction time to system A at this locus is long. ~~The frequency distribution of~~ In a system A which calls for g different genes, each possessing sensors for several other regulatory systems, the reaction time for system A is the maximum distance between the A-sensor and the gene in this set of g genes. The probability distribution of reaction time X has been calculated in terms of the frequency distribution f_s of the number of genes possessing exactly s sensors,

$f_1 = g_1$

on the assumption that ordering of sensors is random. In a homozygous population, for example, the cdf of X is

$$F_X(x) = \prod_r \left(1 - \frac{r}{x+r}\right)^{f_{x+r}}$$

while for heterozygotes

$$H_x(x) = \prod_r \left[1 - \left(\frac{r}{x+r}\right)^2\right]^{f_{x+r}}.$$

This exercise was performed to verify Wallace's conjecture that heterozygotes have a shorter mean reaction time than homozygotes; in fact, we obtain the result that $H_x(x) > F_x(x)$.

COMPARISON OF REACTION TIME DISTRIBUTIONS IN HOMOZYGOUS AND HETEROZYGOUS POPULATIONS.

APPENDICES

D. S. Robson
Cornell University

I. Comparison of reaction time distributions in homozygous and heterozygous populations.

Figures 2 and 3 provide a numerical example illustrating the superiority of heterozygotes over homozygotes in terms of reaction time. The general validity as well as the degree of this superiority may be demonstrated by defining a general form of Figure 2 and then applying the theory of order statistics to obtain the probability distribution of the (discrete) random variable

X = time required for all genes to respond to a particular regulatory system
= time required for the last gene to respond.

Referring to the A-system in Figure 2 we note that this system calls for $g = 5$ different genes and that $f_3 = 3$ of these genes possess 3 sensors and $f_4 = 2$ possess 4 sensors, $f_3 + f_4 = 3 + 2 = 5 = g$. More generally, suppose that the system calls for g different genes among which there are f_1 genes possessing exactly one sensor, f_2 possessing exactly 2, f_3 possessing exactly 3, and so on to the maximum number of sensors, say f_s genes possessing s sensors, with $f_1 + f_2 + f_3 + \cdots + f_s = g$ but with some of the f's possibly equal to zero.

If the ordering of sensors is random at each locus then in a homozygous population the reaction time with respect to the A-system has the (cumulative) probability distribution function

$$F_X(x) \doteq \left(1 - \frac{1}{x+1}\right)^{f_{x+1}} \left(1 - \frac{2}{x+2}\right)^{f_{x+2}} \left(1 - \frac{3}{x+3}\right)^{f_{x+3}} \cdots \left(1 - \frac{s-x}{s}\right)^{f_s} .$$

This function expresses the probability that in a randomly selected homozygote the A-sensor will occur among the first x sensor positions at every one of the g loci which are utilized by the A-system. The average value of X, say μ_F, is then given by

$$\mu_F = s - F_X(1) - F_X(2) - F_X(3) - \cdots - F_X(s-1)$$

which for the A-system in Figure 2 gives

$$4 - \frac{1}{3^3 4^2} - \frac{2^5}{3^3 4^2} - \frac{3^2}{4^2} = 3\frac{13}{36} .$$

Similarly, for the other systems in Figure 2 we obtain:

System	μ_F
A	3 13/16
B	3 25/48
C	3 23/72
D	3 13/16
E	3 25/48
Total	17 1/12

so if these five systems operate in series (in any order) then in the population of homozygotes the average time required for all five systems to act sequentially is exactly 17 1/12 units.

Analagous results for the heterozygous case are obtained by assuming that at each locus the two sensor sequences are independently and randomly ordered. The probability that the nearest A-sensor within pair will occur among the first x sensor positions at every one of the g different loci is then expressed by the cumulative distribution function

$$H_X(x) = \left[1 - \left(\frac{1}{x+1}\right)^2\right]^{f_{x+1}} \left[1 - \left(\frac{2}{x+2}\right)^2\right]^{f_{x+2}} \cdots \left[1 - \left(\frac{s-x}{s}\right)^2\right]^{f_s}$$

and again the mean reaction time is expressed by

$$\mu_H = s - H_X(1) - H_X(2) - \cdots - H_X(s-1)$$

Applied to the example in Figure 2 this gives:

System	μ_H
A	2.682
B	2.817
C	2.614
D	2.682
E	2.817
Total	13.615

compared to a total of 17.083 in the homozygous case.

The general relation $\mu_H < \mu_F$ follows from the relation $H_X(x) > F_X(x)$ for $x = 1,2,3,\ldots,s-1$ $(H_X(s) = F_X(s) = 1)$; i.e., for every x the heterozygote has a larger probability than the homozygote of having a reaction time less than x. This is demonstrated by the relation

$$H_X(x) = \left(1 + \frac{1}{x+1}\right)^{f_{x+1}} \left(1 + \frac{2}{x+2}\right)^{f_{x+2}} \cdots \left(1 + \frac{s-x}{s}\right)^{f_s} F_X(x) \ ;$$

thus,

$$\mu_F - \mu_H = \sum_{x=1}^{s-1} \left[\left(1 + \frac{1}{x+1}\right)^{f_{x+1}} \left(1 + \frac{2}{x+2}\right)^{f_{x+2}} \cdots \left(1 + \frac{s-x}{s}\right)^{f_s} - 1\right] F_X(x) > 0 \ .$$

An indication of the magnitude of this difference may be obtained by considering the least favorable case in which all g genes carry the maximum number of sensors s; i.e., $f_s = g$ and $f_1 = f_2 = \cdots = f_{s-1} = 0$. In this simple case the distribution functions reduce to

$$F_X(x) = \left(\frac{x}{s}\right)^g \qquad H_X(x) = \left[1 - \left(1 - \frac{x}{s}\right)^2\right]^g$$

giving, for example:

g	μ_F	μ_H
1	$\frac{1}{2}s + \frac{1}{2}$	$\frac{1}{3}s + \frac{1}{2} + \frac{1}{6}s^{-1}$
2	$\frac{2}{3}s + \frac{1}{2} - \frac{1}{6}s^{-1}$	$\frac{7}{15}s + \frac{1}{2} + \frac{1}{30}s^{-3}$
3	$\frac{3}{4}s + \frac{1}{2} - \frac{1}{4}s^{-1}$	$\frac{19}{35}s + \frac{1}{2} - \frac{1}{15}s^{-3} + \frac{1}{42}s^{-5}$
4	$\frac{4}{5}s + \frac{1}{2} - \frac{1}{3}s^{-1} + \frac{1}{30}s^{-3}$	$\frac{187}{315}s + \frac{1}{2} - \frac{8}{63}s^{-5} + \frac{1}{30}s^{-7}$
5	$\frac{5}{6}s + \frac{1}{2} - \frac{5}{12}s^{-1} + \frac{1}{12}s^{-3}$	$\frac{437}{693}s + \frac{1}{2} + \frac{8}{63}s^{-5} - \frac{1}{3}s^{-7} + \frac{5}{66}s^{-9}$

Some numerical values of μ_F and μ_H are shown in Table 1 and, in general, these means can be approximated by:

$$\mu_F \doteq \frac{g}{g+1}\, s + \frac{1}{2} \qquad \mu_H \doteq \left[1 - \frac{\Gamma(g+1)\Gamma(\frac{3}{2})}{\Gamma(g+\frac{3}{2})}\right] s + \frac{1}{2} \doteq \left[1 - \frac{\sqrt{\pi}/2}{\sqrt{g+3/4}}\right] s + \frac{1}{2}$$

or, for large s and g,

$$\mu_F - \mu_H \doteq s\left[\frac{.8862269}{\sqrt{g+.75}} - \frac{1}{g+1}\right]$$

We note from Table 1 that the maximum difference occurs in a system of $g = 3$ genes.

Reverse ordering of sequences within a pair has been noted to be more efficient than independent ordering, and in the subpopulation of reverse heterozygotes we note that the cumulative probability distribution say $HR_X(x)$ becomes

$$HR_X(x) = \left(1 - \frac{1}{2x+1}\right)^{f_{2x+1}} \left(1 - \frac{2}{2x+2}\right)^{f_{2x+2}} \cdots \left(1 - \frac{s-2x}{s}\right)^{f_s}$$

This distribution is virtually identical to a homozygous distribution defined on a system with half as many sensors per locus; i.e., in terms of reaction time with respect to any given regulatory gene, this population of reverse heterozygotes has the same statistical properties as a homozygous population with half as many sensors per locus.

Table 1

Comparison between homozygous and heterozygous mean reaction time
with respect to a regulatory system controlling g genes with s sensors per gene

g \ s	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.00*	1.50	2.00	2.50	3.00	3.50	4.00	4.50	5.00	5.50	6.00	6.50	7.00	7.50	8.00
	1.00**	1.25	1.56	1.88	2.20	2.53	2.86	3.19	3.52	3.85	4.18	4.51	4.85	5.18	5.51
2	1.00	1.75	2.44	3.13	3.80	4.47	5.14	5.81	6.48	7.15	7.82	8.49	9.15	9.82	10.49
	1.00	1.44	1.90	2.37	2.83	3.30	3.77	4.23	4.70	5.17	5.63	6.10	6.57	7.03	7.50
3	1.00	1.88	2.67	3.44	4.20	4.96	5.71	6.47	7.22	7.98	8.73	9.48	10.23	10.98	11.73
	1.00	1.58	2.13	2.67	3.21	3.76	4.30	4.89	5.39	5.93	6.47	7.01	7.56	8.10	8.64
4	1.00	1.94	2.79	3.62	4.43	5.24	6.05	6.86	7.66	8.47	9.27	10.07	10.87	11.67	12.48
	1.00	1.68	2.28	2.87	3.47	4.06	4.66	5.25	5.84	6.44	7.03	7.62	8.22	8.81	9.40
5	1.00	1.97	2.86	3.73	4.58	5.43	6.27	7.11	7.95	8.79	9.63	10.47	11.30	12.14	12.97
	1.00	1.76	2.39	3.02	3.65	4.28	4.91	5.54	6.18	6.81	7.44	8.07	8.70	9.33	9.96
6	1.00	1.98	2.91	3.81	4.69	5.56	6.43	7.29	8.16	9.02	9.88	10.74	11.60	12.46	13.32
	1.00	1.82	2.48	3.14	3.80	4.45	5.11	5.77	6.43	7.09	7.75	8.41	9.07	9.73	10.39
7	1.00	1.99	2.94	3.86	4.76	5.65	6.54	7.43	8.31	9.19	10.07	10.95	11.83	12.71	13.59
	1.00	1.87	2.55	3.23	3.91	4.59	5.27	5.95	6.64	7.32	8.00	8.70	9.36	10.04	10.73
8	1.00	2.00	2.96	3.90	4.81	5.72	6.63	7.53	8.43	9.32	10.22	11.11	12.00	12.90	13.79
	1.00	1.90	2.60	3.30	4.00	4.70	5.40	6.10	6.80	7.50	8.21	8.91	9.61	10.31	11.01
9	1.00	2.00	2.97	3.92	4.86	5.78	6.69	7.61	8.52	9.43	10.33	11.24	12.14	13.05	13.95
	1.00	1.92	2.65	3.36	4.08	4.80	5.51	6.23	6.95	7.66	8.38	9.09	9.81	10.53	11.24
10	1.00	2.00	2.98	3.94	4.89	5.82	6.75	7.67	8.59	9.51	10.42	11.34	12.25	13.17	14.08
	1.00	1.94	2.69	3.42	4.15	4.88	5.61	6.34	7.07	7.80	8.53	9.26	9.99	10.72	11.45
11	1.00	2.00	2.99	3.96	4.91	5.85	6.79	7.72	8.65	9.58	10.50	11.42	12.35	13.27	14.19
	1.00	1.96	2.72	3.47	4.21	4.95	5.69	6.43	7.17	7.91	8.66	9.40	10.14	10.88	11.62

*homozygous μ_F **heterozygous μ_H